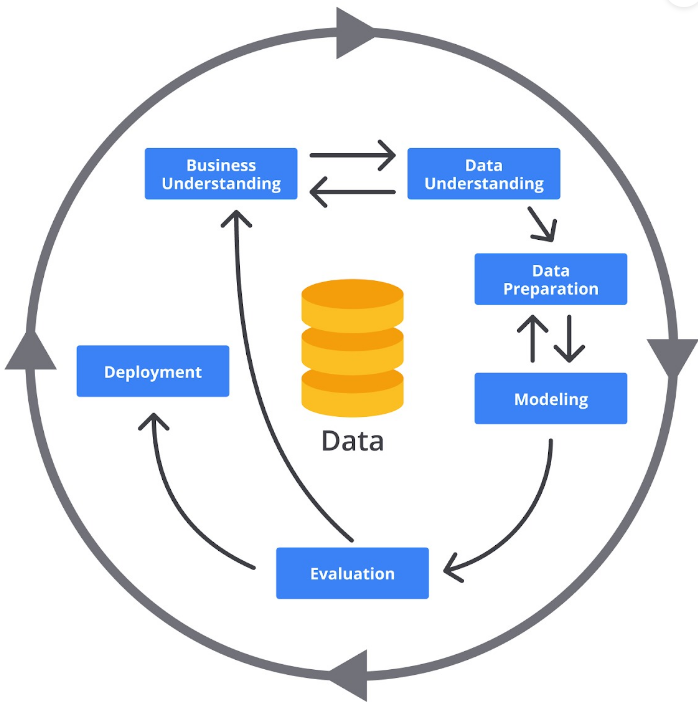
**Metodologi Data Science**

**CRISP-DM**



CRISP-DM atau *Cross Industry Standard Process for Data Mining*adalah proses model yang berfungsi sebagai dasar untuk proses data science. Proses model CRISP-DM memberikan gambaran tentang siklus hidup proyek data mining yang terdapat pada big data.

Sekilas info, *data mining*berarti proses pengumpulan informasi penting dari data dalam jumlah yang besar. CRISP-DM memiliki enam tahapan yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation,*dan *Deployment*.

1. Business Understanding

Pemahaman mendalam tentang kebutuhan pelanggan. Menentukan tujuan bisnis, menilai situasi ketersediaan sumber daya, tujuan pengumpulan data, menghasilkan rencana proyek.

1. Data Understanding

Fokus untuk mengidentifikasi, mengumpulkan, dan menganalisis Kumpulan data yang dapat membantu mencapai tujuan proyek. Mengumpulkan data, menjelaskan data, jelajahi data, dan verfifikasi kualitas data.

1. Data Preparation

Disebut “data mining”, yaitu menyiapkan Kumpulan data akhir untuk permodelan. Kegiatan pada fase ini diantaranya memilih data yaitu menentukan set data yang akan digunakan, lalu bersihkan data, bangun data atau dengan karta lain mendapatkan atribu baru yang akan membantu, lalu integrasikan data dan yang terakhir format data.]

1. Modeling

Membuat dan menilai berbagai model berdasarkan beberapa Teknik permodelan yang berbeda. Ada 4 tugas, yaitu Teknik permodelan, menghasilkan desain pengujian, membangun model, dan yang terakhir menilai model.

1. Evaluation

Melihat lebih luas model yang paling sesuai bisnis dan yang harus dilakukan selanjutnya.

Terdapat 3 kegiatan, yaitu evaluasi hasil, proses peninjauan, dan menentukan Langkah selanjutnya.

1. Deployment

Perencanaan untuk deployment dimulai dari fase *business understanding*, dan harus menggabungkan tidak hanya untuk menghasilkan nilai model, tetapi juga cara mengonversi skor keputusan, dan cara untuk menggabungkan Keputusan tersebut.